

HW02-价量多因子模型

22335047 郭振宇

2024 年 5 月 23 日

摘要: 这篇报告着手构建基于价量特征的多因子模型，以提高股票选择的性能。首先，我们选择了国泰君安研报中的 20 个价量因子，并根据开源证券的研报复现了其中的三个价量因子。接着，利用有效性分析函数对这些因子进行有效性识别，并筛选出有效因子。最后，基于选出的因子，通过历史 IC 值加权的方式构建 Alpha 因子，并进行因子组合的有效性分析。通过对 Alpha 因子的数据观察，发现该因子在长期投资中可能具有一定的价值，但其短期内的波动和不稳定性需要投资者特别关注。

关键词: 价量因子、多因子模型、有效性分析、Alpha 合成

1 问题描述

参考国泰君安《基于短周期价量特征的多因子选股体系》，尝试构造多个价量因子（至少包括该研报中 20 个价量因子以及 3 个以上原创因子（可复现数据文档中其他三篇研报）），需给出部分研报中价量因子的描述和分析，以及所有原创因子的描述和分析。最后可通过线性模型或机器学习模型进行因子结合成最后的 alpha，并给出 alpha 的有效性分析或回测报告。

2 因子收集与构造

2.1 国泰君安研报中的 20 个价量因子

1. 价量背离因子，相关系数越低，后期实现超额收益的概率越高，相关结果见 HW01。

$$-1 * CORR(VAWP, VOLUME, d) = -corr(vwap_{t-d:t}^i, volume_{t-d:t}^i) \quad (1)$$

2. 开盘缺口因子，当日个股开盘缺口，短周期内有较强的动量效应。

$$OPEN/DELAY(CLOSE, 1) = \frac{open_t^i}{close_{t-1}^i} \quad (2)$$

3. 异常成交量因子，成交量的异常变化往往预示着阶段性反转，如异常缩量是阶段底部的特征。

$$-1 * VOLUME/MEAN(VOLUME, d) = -\frac{volume_t^i}{mean(volume_{t-d:t}^i)} \quad (3)$$

4. 量幅背离因子，相关系数越低，后期实现超额收益的概率越高。

$$-1 * CORR(HIGH/LOW, VOLUME, d) = -corr(\frac{high_{t-d:t}^i}{low_{t-d:t}^i}, volume_{t-d:t}^i) \quad (4)$$

有效性分析结果如Figure 1，从图中可以看出，IC 值随日期变化呈现出明显的波动性，在某些日期震荡剧烈，整体 IC 均值为 0.08，说明该因子能够有效预测收益率。因子收益率和累计因子收益率呈现出一定的波动性。在某些时期，两者呈现正相关，在某些时期则呈现负相关。

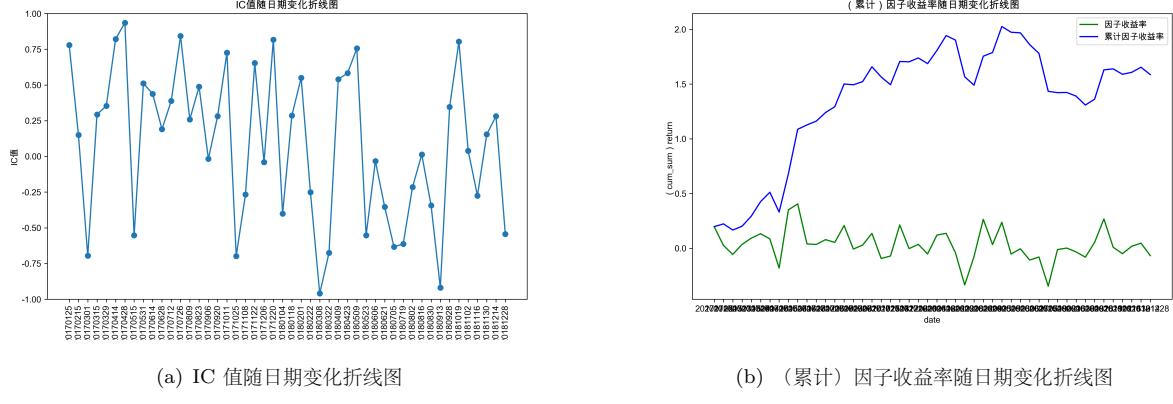


图 1: 量幅背离因子的有效性分析

5. 该因子计算股票价格在高低区间的相对位置变化率，并取一阶差分的负值来衡量价格变动的趋势。

$$\begin{aligned} Alpha2 : & -1 * DELTA(((CLOSE - LOW) - (HIGH - CLOSE)) / (HIGH - LOW)), 1) \\ & = -\left(\frac{(close_t^i - low_t^i) - (high_t^i - close_t^i)}{high_t^i - low_t^i} - \frac{(close_{t-1}^i - low_{t-1}^i) - (high_{t-1}^i - close_{t-1}^i)}{high_{t-1}^i - low_{t-1}^i}\right) \quad (5) \end{aligned}$$

6. 该因子通过计算最高价与最低价的几何平均值减去成交量加权平均价，来衡量价格的偏离程度。

$$Alpha13 : (((HIGH * LOW)^{0.5}) - VWAP) = (high_t^i * low_t^i)^{0.5} - vwap_t^i \quad (6)$$

7. 该因子衡量股票的短期收益率，正值表明股价在五天内上升，负值则表明股价下降。

$$Alpha14 : CLOSE - DELAY(CLOSE, 5) = close_t^i - close_{t-5}^i \quad (7)$$

8. 该因子通过计算过去 12 天收盘价均值与当前收盘价的比值来衡量价格的相对水平。比值大于 1 表明当前价格低于平均水平，可能存在低估；比值小于 1 表明当前价格高于平均水平，可能存在高估。

$$Alpha34 : MEAN(CLOSE, 12) / CLOSE = \frac{\text{mean}(close_{t-11:t}^i)}{close_t^i} \quad (8)$$

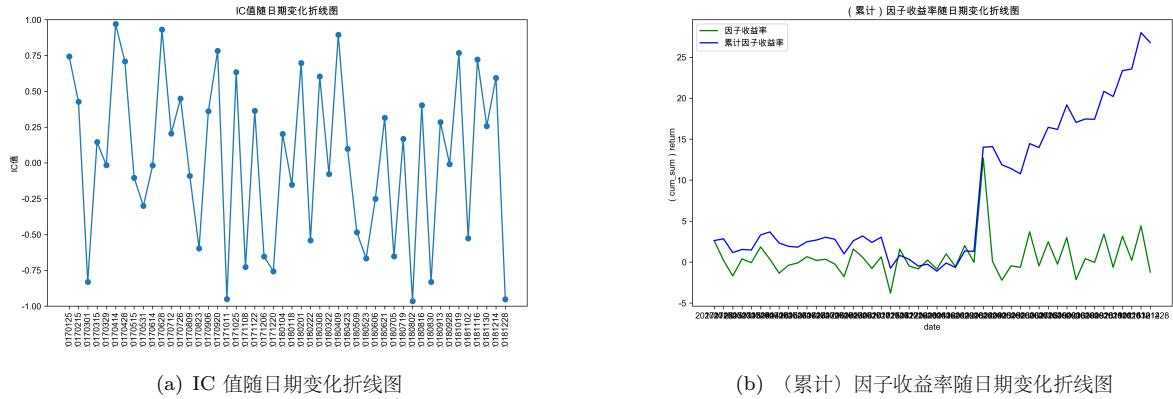


图 2: Alpha34 的有效性分析

有效性分析结果如Figure 2，整体的 IC 均值高于 0.03，说明该因子能够有效预测未来收益率。从图中可以看到，因子收益率在某些时期出现较大的涨跌幅，这可能是由于该因子的算法对市场短期波动比较敏感所致。与因子收益率相比，累计因子收益率增长要快得多，说明该因子在长期投资中表现较好。虽然整体趋势向好，但也出现了几次较大幅度的回撤，所以仍需关注其短期波动性和回撤风险。

9. 该因子反映了成交量与价格之间的动态关系及其短期变化，高值表示近期成交量与价格的相关性较强，可能预示着市场趋势的延续；低值则表示相关性较弱，可能预示着市场趋势的反转。相关系数衡量两个变量的线性关系，排名和求和则强调了短期内这种关系的强度和持续性。

$$\begin{aligned} Alpha36 : & RANK(SUM(CORR(RANK(VOLUME), RANK(VWAP)), 6), 2) \\ & = rank(\sum(corr(rank(volume_{t'-9:t'}^i), rank(vwap_{t'-9:t'}^i))_{t-5:t}, 2)) \end{aligned} \quad (9)$$

10. 该因子衡量最高价和成交量排名之间在过去 5 天的相关性。负相关系数表明，当成交量排名上升时，最高价可能下降，反之亦然，该因子可以用于捕捉市场反转信号。

$$\begin{aligned} Alpha62 : & -1 * CORR(HIGH, RANK(VOLUME), 5) \\ & = -corr(high_{t-4:t}^i, rank(volume_{t-4:t}^i)) \end{aligned} \quad (10)$$

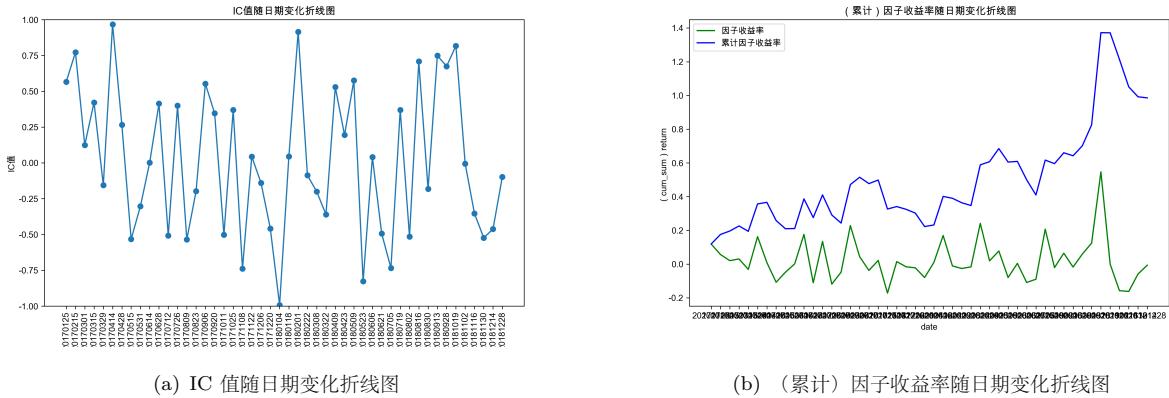


图 3: Alpha62 的有效性分析

有效性分析结果如Figure 3，该因子在不同时间段内表现出较大的波动性，从因子收益率和 IC 值的波动可以看出，短期内的预测能力不稳定。累积因子收益率虽然整体上升，但中间存在显著的回撤风险。这些分析表明，该因子可能在一定市场环境下有效，但其整体稳定性和一致性较差。

11. 该因子衡量股票价格相对于其 24 天均值的偏离程度，正值表示当前价格高于 24 天均值，表明短期内价格上升趋势；负值则表示当前价格低于 24 天均值，表明短期内价格下降趋势。

$$\begin{aligned} Alpha71 : & (CLOSE - MEAN(CLOSE, 24))/MEAN(CLOSE, 24) * 100 \\ & = \frac{close_t^i - mean(close_{t-23:t}^i)}{mean(close_{t-23:t}^i)} * 100 \end{aligned} \quad (11)$$

12. 该因子首先对最高价和成交量进行排名，然后计算它们在过去 5 天内的协方差，并对结果进行排名，最后取负值。协方差衡量两个变量的共同变动趋势，负值表示反向变动。

该因子通过结合价格与成交量的关系，捕捉市场上潜在的价格反转或动量信号。其值越低，表明短期内价格可能存在较大波动，适用于短期交易策略。

$$\begin{aligned} Alpha88 : & (-1 * RANK(COVIANCE(RANK(HIGH), RANK(VOLUME), 5))) \\ & = -rank(covariance(rank(high_{t'-9:t'}^i), rank(volume_{t'-9:t'}^i))_{t-4:t}) \end{aligned} \quad (12)$$

13. 该因子通过计算过去 20 天的交易金额的标准差来衡量市场的波动性。标准差越大，表示交易金额波动越剧烈，市场不确定性较高；标准差越小，表示交易较为稳定，市场波动性较低。

$$Alpha95 : STD(AMOUNT, 20) = std(amount_{t-19:t}^i) \quad (13)$$

有效性分析结果如Figure 4，整体的 IC 均值为-0.027，说明该因子预测的方向与实际方向相反，累积因子收益率曲线也呈现出相应的现象。根据因子的定义可知，因子的值越小，表明交易金额波动越小，市场波动性越小，相应地未来收益率就会越稳定，反之亦然。所以，该因子是个反向预测因子。

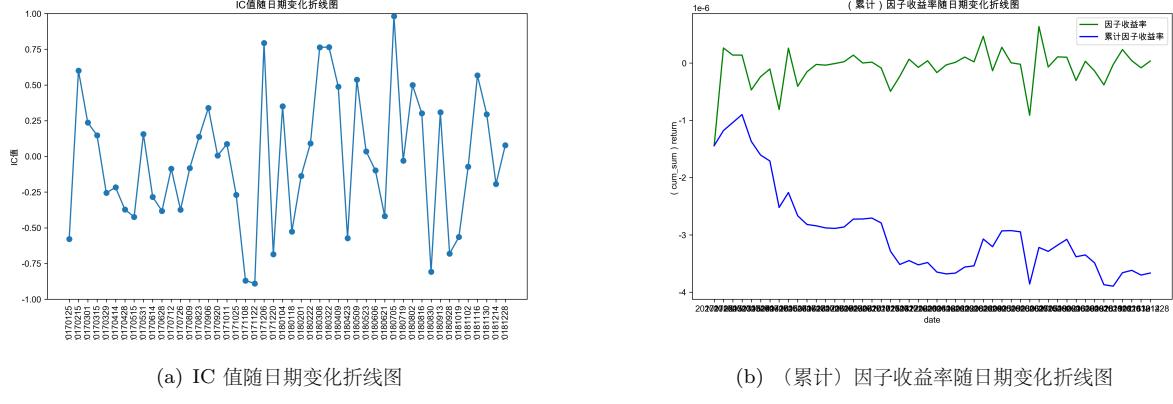


图 4: Alpha95 的有效性分析

14. 该因子通过计算开盘价和成交量的排名相关系数，并取其负值来反映市场行为。负的相关系数表明开盘价和成交量之间存在反向关系，当相关系数为负且绝对值较大时，可能预示着市场将发生反转。

$$\begin{aligned} \text{Alpha105} : & -1 * \text{CORR}(\text{RANK}(OPEN), \text{RANK}(VOLUME), 10) \\ & = -\text{corr}(\text{rank}(\text{open}_{t-9:t}^i), \text{rank}(\text{volume}_{t-9:t}^i)) \end{aligned} \quad (14)$$

15. 该因子反映了股票的典型价格，综合了收盘价、最高价和最低价的信息，可以用于判断股票的平均交易水平，有助于识别价格的中枢位置。

$$\text{Alpha126} : (\text{CLOSE} + \text{HIGH} + \text{LOW}) / 3 = (\text{close}_t^i + \text{high}_t^i + \text{low}_t^i) / 3 \quad (15)$$

16. 该因子通过计算当前收盘价与 12 天前收盘价的变化率，并乘以成交量来衡量价格变化的力度。价格变化幅度越大且成交量越高，因子值越大，表明市场对价格变动的确认度越高。

$$\begin{aligned} \text{Alpha135} : & (\text{CLOSE} - \text{DELAY}(\text{CLOSE}, 12)) / \text{DELAY}(\text{CLOSE}, 12) * \text{VOLUME} \\ & = \frac{\text{close}_t^i - \text{close}_{t-12}^i}{\text{close}_{t-12}^i} * \text{volume}_t^i \end{aligned} \quad (16)$$

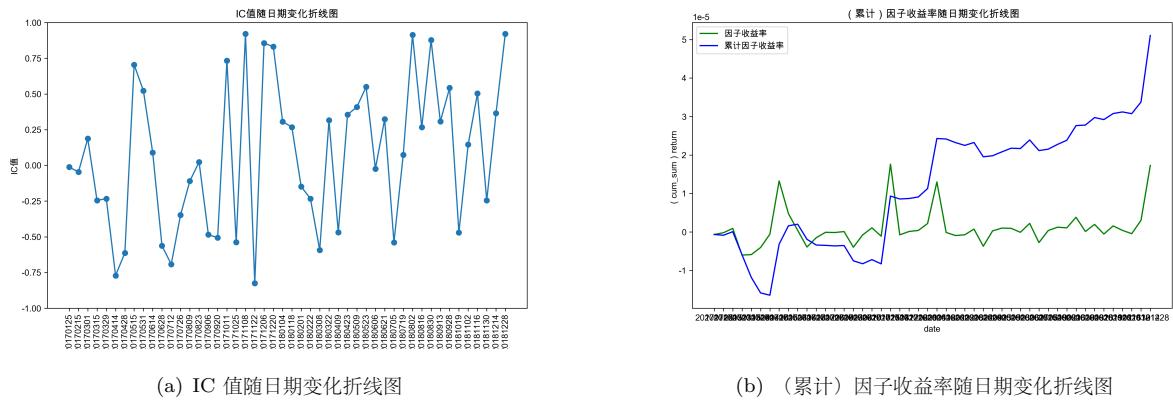


图 5: Alpha135 的有效性分析

因子有效性分析结果如Figure 5，整体的 IC 均值高于 0.03，说明该因子能够有效预测未来收益率。因子收益率的波动性较大，说明因子在不同时间段的表现不尽相同；累计因子收益率长期来看是正的，表明因子在总体上是有效的，能够带来正收益。总体来看，该因子具有一定的有效性，能够在较长时间内带来正的累计收益，但其短期表现波动较大。

17. 该因子衡量开盘价和成交量在过去 10 天内的相关性，负相关系数表示开盘价和成交量之间存在反向关系，可能预示着市场的趋势反转。

$$\text{Alpha139} : -1 * \text{CORR}(\text{OPEN}, \text{VOLUME}, 10) = -\text{corr}(\text{open}_{t-9:t}^i, \text{volume}_{t-9:t}^i) \quad (17)$$

18. 该因子通过计算短期（9 天）与长期（26 天）的成交量均值的差异，并以 12 天的均值为基准进行标准化，来衡量成交量的变化趋势，有助于判断市场的活跃程度和潜在的价格变动方向。

$$\begin{aligned} \text{Alpha151} &: (\text{MEAN}(\text{VOLUME}, 9) - \text{MEAN}(\text{VOLUME}, 26)) / \text{MEAN}(\text{VOLUME}, 12) * 100 \\ &= \frac{\text{mean}(\text{volume}_{t-8:t}^i) - \text{mean}(\text{volume}_{t-25:t}^i)}{\text{mean}(\text{volume}_{t-11:t}^i)} * 100 \end{aligned} \quad (18)$$

19. 该因子反映了股票在过去 1 天内的价格变动幅度，并结合成交量来衡量变化的力度。

$$\begin{aligned} \text{Alpha178} &: (\text{CLOSE} - \text{DELAY}(\text{CLOSE}, 1)) / \text{DELAY}(\text{CLOSE}, 1) * \text{VOLUME} \\ &= \frac{\text{close}_t^i - \text{close}_{t-1}^i}{\text{close}_{t-1}^i} * \text{volume}_t^i \end{aligned} \quad (19)$$

20. 该因子通过结合成交量均值与最低价的相关性以及价格中枢来衡量市场偏离度。可以用于评估当前价格相对于市场均值的偏离程度，有助于判断市场是否存在高估或低估。

$$\begin{aligned} \text{Alpha191} &: (\text{CORR}(\text{MEAN}(\text{VOLUME}, 20), \text{LOW}, 5) + ((\text{HIGH} + \text{LOW}) / 2)) - \text{CLOSE} \\ &= \text{corr}(\text{mean}(\text{volume}_{t'-19:t'}^i)_{t-4:t}, \text{low}_{t-4:t}^i) + ((\text{high}_t^i + \text{low}_t^i) / 2) - \text{close}_t^i \end{aligned} \quad (20)$$

2.2 复现的 3 个价量因子

2.2.1 开源证券-长端动量因子

1. 因子构建方式

(1) 对选定股票，回溯取其最近 160 个交易日的数据；(2) 计算股票每日的振幅（最高价/最低价-1）；(3) 选择振幅较低的 70% 交易日，涨跌幅加总。

2. 因子计算公式

$$\text{Alpha1} = \text{sum}(\text{pct_chg}_t^i, (\text{rank}((\text{high}_t^i / \text{low}_t^i - 1), 160), 0.7 * 160)) \quad (21)$$

3. 因子的有效性分析

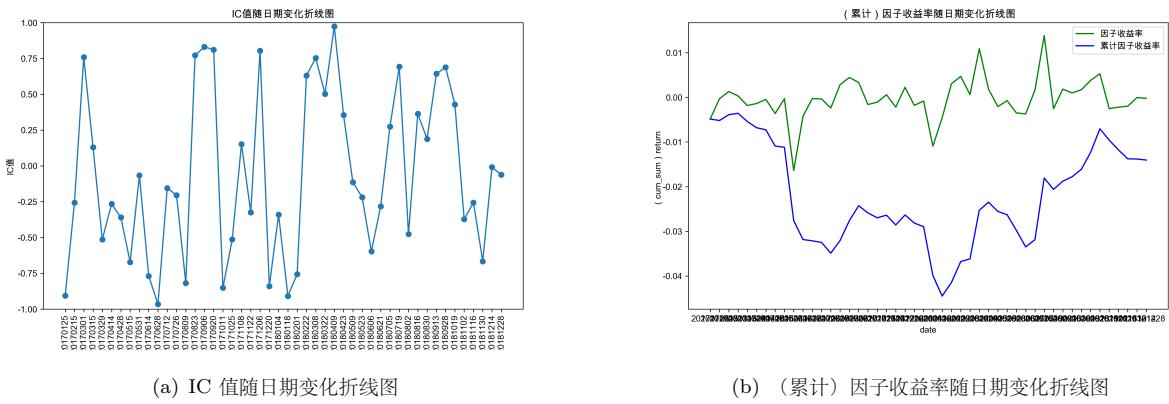


图 6: 因子有效性分析结果

该因子能够捕捉低波动性环境下的收益特征，低振幅通常意味着市场波动较小，价格走势较为平稳。在这种环境下，股票的涨跌幅可能更能反映市场的真实趋势，而非由短期波动或市场噪音驱动。因此，选取振幅较低的 70% 交易日的涨跌幅进行加总，可以帮助识别出在稳定市场条件下的平均收益表现。

有效性分析结果如Figure 6，该因子整体的 IC 均值高于 0.03，说明该因子能够有效预测未来收益率。从 IC 值的波动来看，因子的预测能力并不稳定，可能受到市场环境、股票本身波动等多种因素的影响；从累计因子收益率来看，整体表现并不理想，尤其是在某些时间段内出现了明显的回撤；因子在不同时间段表现差异较大，需要进一步优化因子构建方法，例如调整振幅计算方法，选择不同的参数（如测试时间、选取比例等），或者结合其他因子进行多因子策略。

2.2.2 开源证券-理想反转因子

1. 因子构建方式

(1) 对选定股票 S，回溯取其过去 20 个交易日的数据；(2) 计算股票 S 每日的平均单笔成交金额；(3) 平均单笔金额高的 10 个交易日，涨跌幅加总，记作 M_{high} ；平均单笔金额低的 10 个交易日，涨跌幅加总，记作 M_{low} ；(4) 计算 $M = M_{high} - M_{low}$ 。

2. 因子计算公式

$$\begin{aligned} M_{high} &= \text{sum}(pct_chg_t^i, (\text{rank}(vwap^i, 20), 0.1 * 20)) \\ M_{low} &= \text{sum}(pct_chg_t^i, (\text{rank}(-vwap^i, 20), 0.1 * 20)) \\ M &= M_{high} - M_{low} \end{aligned} \quad (22)$$

3. 因子描述与分析

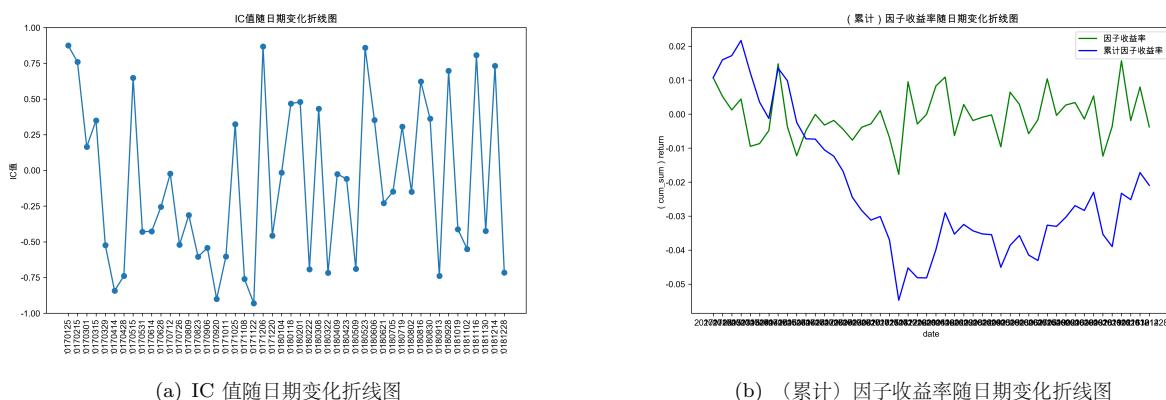


图 7: 因子有效性分析结果

该因子通过比较高和低平均单笔成交金额交易日的涨跌幅，来捕捉市场中可能存在的反转现象。理想反转因子的正负值可以揭示市场中的反转信号：(1) M 为正，表明平均单笔成交金额高的交易日涨幅较大，而低的交易日涨幅较小，可能预示着市场在大资金参与下有上涨的潜力；(2) M 为负，表明平均单笔成交金额低的交易日涨幅较大，而高的交易日涨幅较小，可能预示着市场在大资金影响下有回调的风险。

有效性分析结果如Figure 7，整体的 IC 均值高于 0.03，但由于因子收益率和累积因子收益率存在显著波动，且 IC 值在正负之间频繁交替，所以因子在整体上表现出非一致性和不稳定性。可能是因为该因子在选择的股票样本上表现不佳，可以考虑增加股票数据样本量，或结合其他因子来增强预测效果。

2.2.3 开源证券-理想振幅因子

1. 因子构建方式

(1) 对选定股票 S，回溯取其过去 20 个交易日的数据；(2) 计算股票 S 每日的振幅（最高价/最低价-1）；(3) 选择收盘价较高的 $\lambda(25\%)$ 有效交易日，计算振幅均值得到高价振幅因子 V_{high} ；选择收盘价较低的 $\lambda(25\%)$ 有效交易日，计算振幅均值得到低价振幅因子 V_{low} ；(4) 计算 $V = V_{high} - V_{low}$ 。

2. 因子计算公式

$$\begin{aligned}
 V_{high}(\lambda) &= \text{mean}(high_t^i / low_t^i - 1, (\text{rank}(close^i, 20), \lambda * 20)) \\
 V_{low}(\lambda) &= \text{mean}(low_t^i / high_t^i - 1, (\text{rank}(-close^i, 20), \lambda * 20)) \\
 V &= V_{high}(\lambda) - V_{low}(\lambda)
 \end{aligned} \tag{23}$$

3. 因子描述与分析

该因子通过比较高收盘价和低收盘价交易日的振幅差异，来捕捉市场中可能存在的波动性特征。理想振幅因子的正负可以揭示市场的波动特征：(1) V 为正，表明高收盘价交易日的振幅较大，低收盘价交易日的振幅较小，可能预示着市场在上涨趋势中波动性较高，而在下跌趋势中波动性较低；(2) V 为负，表明低收盘价交易日的振幅较大，高收盘价交易日的振幅较小，可能预示着市场在下跌趋势中波动性较高，而在上涨趋势中波动性较低。

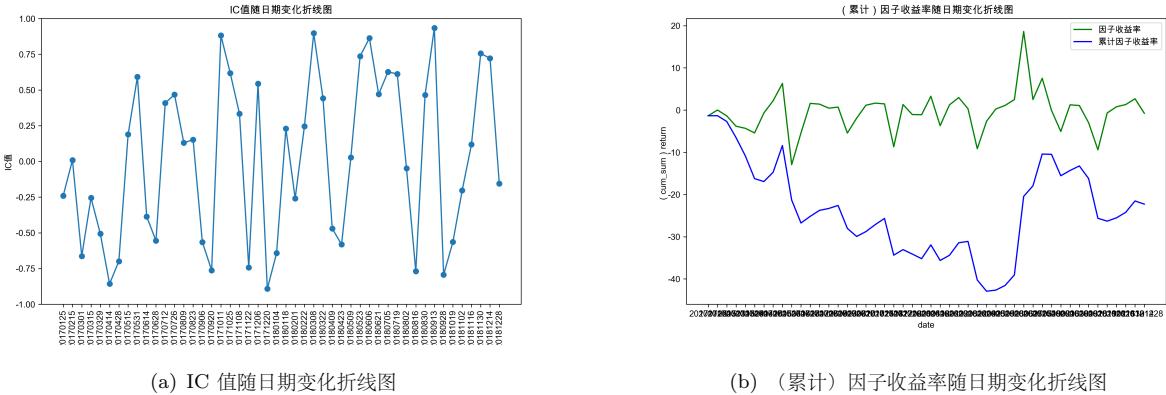


图 8: 因子有效性分析结果

有效性分析结果如Figure 8，从 IC 值的波动来看，因子的预测能力并不稳定，因子收益率的波动性较大且累计收益率表现不佳，表明该因子在长期持有策略下可能并不具备稳定的盈利能力。原因可能是选取的股票样本数较少，后续可以考虑增加股票数据样本量，或结合其他因子进行多因子策略。

3 因子描述与分析

3.1 数据描述

本报告的数据来自于 Tushare，Tushare 是一个开源的金融数据接口，能够提供股票、期货、外汇等金融数据，并提供数据清洗、分析、回测等功能。

有效性分析采用的数据集为 A 股上市公司 2016 年 12 月 31 日至 2018 年 12 月 31 日的交易数据。数据包括股票代码、日期、开盘价、最高价、最低价、收盘价、成交量、成交金额、换手率等信息。但由于 Tushare 接口对调取范围以及访问频率的限制，本报告只选取了部分股票进行相关研究和分析。

3.2 因子有效性分析

通过 HW01 的因子有效性分析函数，可以得到如下Table 1的 23 个因子的相关数据，包括 IC 均值、ICIR 值、负相关 IC 值占比以及回归 t 检验均值（其中两个 NAN 值表示因子无法拟合线性回归模型）。国泰君安研报中的部分价量因子以及复现的开源证券研报中的 3 个因子的详细分析如上所述。

表 1: alpha 有效性分析数据

因子	IC 均值	ICIR 值	负相关 IC 值占比	回归 t 检验均值
1	0.136504	0.25189	0.375	0.545724
2	-0.000968	-0.001659	0.5	-0.015842
3	0.027099	0.051316	0.5	0.115322
4	0.088241	0.167169	0.416667	0.182217
5	-0.023153	-0.048838	0.5625	0.05498
6	0.059733	0.115882	0.458333	0.179574
7	0.054904	0.091115	0.458333	0.33027
8	0.032366	0.055938	0.4375	0.054289
9	-0.233333	-0.539761	0.541667	NAN
10	0.019767	0.038946	0.5	-0.05336
11	0.027663	0.050868	0.5	0.192946
12	-0.074859	-0.154787	0.479167	NAN
13	-0.027246	-0.058643	0.520833	0.041837
14	0.039386	0.072475	0.416667	0.152664
15	0.060215	0.117619	0.4375	0.191924
16	0.074945	0.147811	0.458333	0.32919
17	0.038443	0.074724	0.4375	0.106224
18	-0.011122	-0.02164	0.4375	0.076953
19	0.012796	0.024832	0.5	0.021523
20	-0.048079	-0.092259	0.479167	-0.098371
21	-0.058184	-0.101134	0.604167	-0.162635
22	-0.09062	-0.160745	0.625	-0.225411
23	0.017605	0.031133	0.458333	0.094181

3.3 Alpha 因子合成

3.3.1 因子合成方法

因子合成时，将共线性比较严重的因子先进行合成，再进行多元回归，可提升回归问题的准确性。常见的因子合成方法有：等权法；因子收益率（半衰）加权法；因子 IC（半衰）加权法；最大化 IC_IR 法；最大化 IC 法和主成分分析法。

本报告的因子合成方法是基于 IC 值的加权平均法，具体步骤如下：(1) 根据有效性分析函数，计算各 alpha 的 IC 值；(2) 对 IC 值进行归一化，计算各 alpha 的权重系数；(3) 加权平均得到最终的 Alpha。

3.3.2 alpha 因子的有效性分析

将得到的 Alpha 代入到 HW01 的因子有效性分析函数中，得到相应的数据和图表。

表 2: Alpha 因子有效性分析数据

因子	IC 均值	ICIR 值	负相关 IC 值占比	回归 t 检验均值
Alpha	0.033224	0.073045	0.479167	0.114775

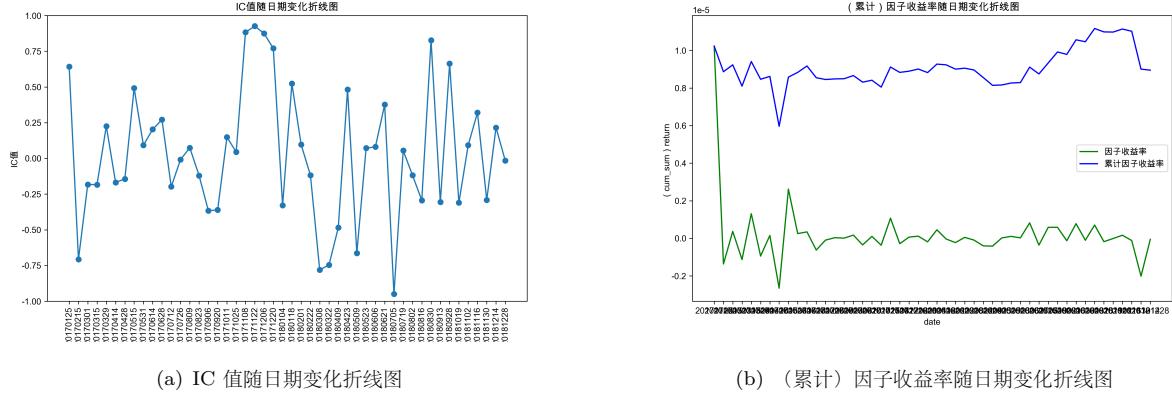


图 9: Alpha 有效性分析结果

有效性分析结果如Figure 9, 根据表Table 2和9(a)可知 IC 均值高于 0.03, 说明 Alpha 因子的预测能力较强, 同时也表明 Alpha 的预测性能缺乏一致性和稳定性; 根据9(b)可以观察到, 因子收益率和累计因子收益率的波动性在长期持有策略下基本保持稳定, 但在某些时间段也有明显的回撤, 这可能和市场环境、选择的股票样本、因子的选择以及加权方式有关。总的来说, 合成的 Alpha 能够在长期保持稳定的情况下, 对市场的走势进行预测, 但其预测能力仍然有待提高。

4 不足之处与展望

此报告的主要内容是基于 A 股上市公司 2016 年 12 月 31 日至 2018 年 12 月 31 日的交易数据, 对研报中的价量单因子进行有效性分析, 筛选出部分因子后构建合成 Alpha 因子, 以此构建价量多因子模型并进行分析。上述研究目标和研究流程已经基本上实现, 但是报告中存在以下局限:

1. 由于 Tushare 接口的限制, 只能选取部分股票进行因子分析, 导致回测的样本量太少, 因子的有效性分析结果呈现出非一致性和不稳定性, 且因子的选取和筛选依赖于主观属性。
2. 因子有效性分析方面, 仅对信息系数 (IC) 值进行了简单分析, 缺乏对因子有效性的深入验证。实际应用中, 可以采用不同的有效性评估指标, 在不同的时间区间和样本中进行检验等。
3. 因子合成方法仅基于历史 IC 值的加权平均法, 没有考虑其他合成方法, 如基于机器学习的因子合成方法等。实际应用中, 需要对其他因子合成方法进行深入研究和比较, 以找出最优的合成策略。

总而言之, 在实际应用中, 需要进一步完善因子分析的有效性验证, 并探索更多样化的因子合成方法, 以提高因子投资组合的收益和风险管理能力。

参考文献

- [1] 国泰君安, 基于短周期价量特征的多因子选股体系 – 数量化专题之九十三, 证券研究报告, 2017.06.15.
- [2] 开源证券, 独家量化因子的高频测试, 证券研究报告, 2021.09.24.

附录

本报告的所有代码见附件, 也同步上传到了 GitHub 个人主页: <https://github.com/zeyyo-guo>